

Data Management Systems

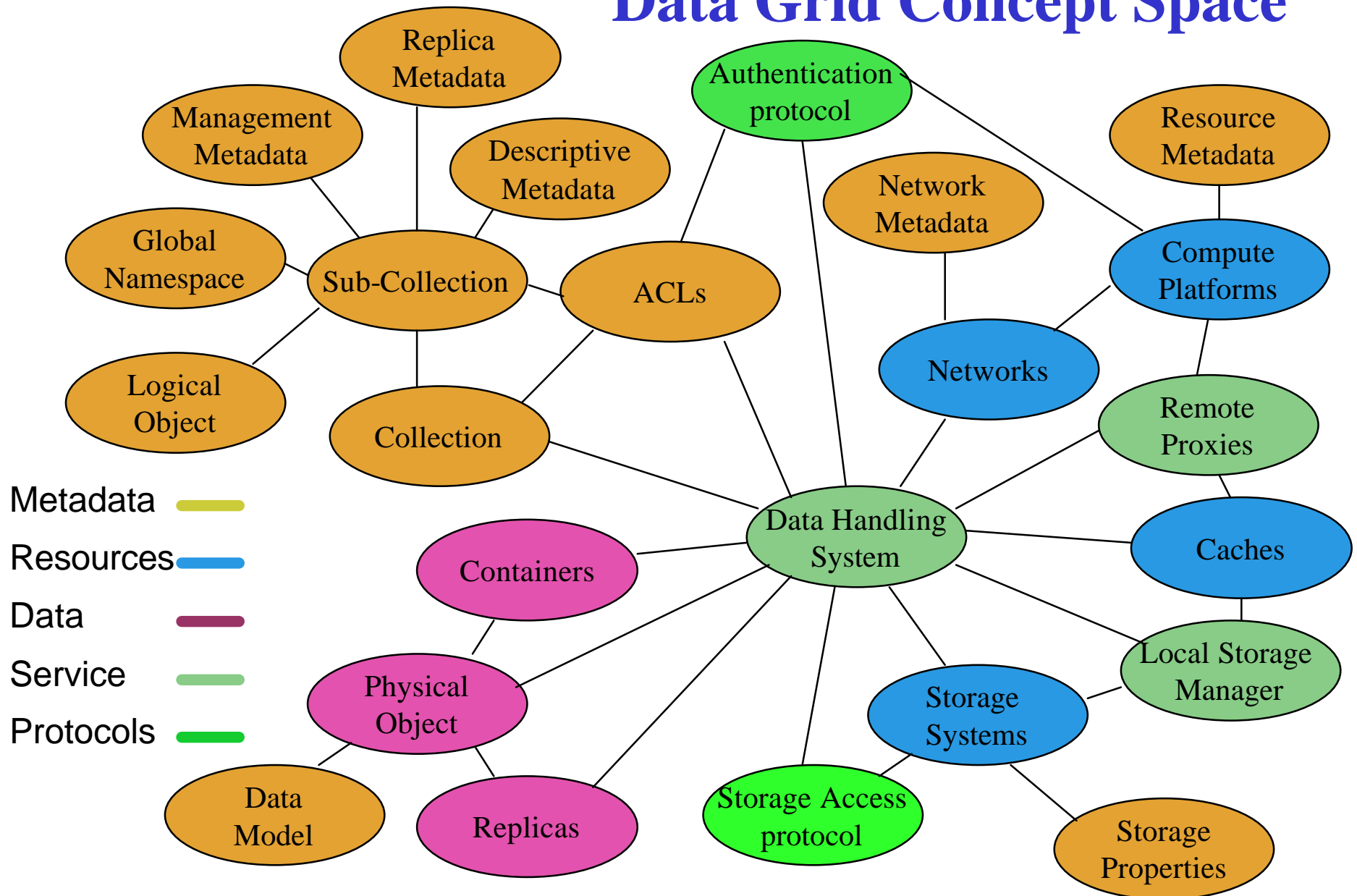
Global Grid Forum 2
June 16, 2001



Data Dictionary

- Define Protocols
 - Arguments
 - Command sequence
- Define APIs
 - Parameters
- Services
 - Functionality
- SDKs - Software Development Kits
 - Capabilities

Data Grid Concept Space



Are There Multiple Grids?

- Grid - integration of physical resources to support execution environment
- Service grid - integration of multiple service environments (OMG MDA, Digital Libraries)
- Data Grid - management of persistent data entities across multiple repositories (OMG CMW)
- Virtual Data Grid - integration of data grid and the Grid

URLs

- Darpa Agent Markup Language, DAML-S
 - <http://www.daml.org>
- OMG Model Driven Architecture
 - <http://www.omg.org/mda>
- OMG Common Warehouse Metamodel
 - <http://www.omg.org/cwm>

Data Grids

- Data Grids provide mechanisms to control data entities:
 - Persistence
 - Latency management
 - Global logical name space
 - Advanced naming mechanisms (collection-based, knowledge-based, “wisdom-based”)

Collections

Collection - managed data

Use database to organize attributes about data objects

Separate information management from data storage

Support APIs for information discovery, data access



Integration accomplished through a data handling system that characterizes the storage systems

Distributed Collections

Distributed Data Collection

Same name space

Same schema

Separate administration domains

Heterogeneous database instances



Integration requires the ability to characterize both the schemas and the table structures of each information repository

Data Grids

Data Grid - linking multiple data collections

Separate name spaces

Separate schema

Separate administration domains

Heterogeneous database instances



The data grid is itself a collection that provides mechanisms to hide latency and manage semantics

Virtual Data Grid

Link multiple data collections and provide ability to execute processes to recreate derived data

- Separate name spaces
- Latency management
- Staging from collection into containers
- Manipulation of objects in containers



The virtual data grid integrates collection management and grid protocols to process data

Persistent Archive

Persistent archive

- Describe archived data as collections
- Describe processes used to create collections
- Manage evolution of technology



The persistent archive is itself a virtual data grid that provides mechanisms to manage relationships over time

Interfaces

- Is data management in a grid relative to the execution environment rather than the persistence environment?
- Is the interface between Collections and the Grid a staging interface?
 - Stage data from HRMs
 - Stage data into containers from a remote collection
 - Generalized staging interface between persistent data management environment and grid, used by Virtual Data Grid